the combined similarity information, an objective function is iteratively maximized in order to yield a generalized similarity value that expresses the similarity of particular pairs of documents. In an embodiment, the generalized similarity value is used to determine the proper category, among a taxonomy of categories in an index, cache or search system, into which certain documents belong.

In the Claims:

In the claims, please amend claims 1, 19, 20, 21, 28, 29 and 34. Please delete claims 35 – 37.

1. (Amended) A method of categorizing a plurality of new electronic documents into a set of categories, comprising the steps of:

establishing a plurality of training sets, wherein each training set is associated with a category and includes training documents that have been classified as belonging to said associated category;

determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories by determining similarity between said each document and the training documents that belong to the training set of said category; and

wherein the step of determining similarity is performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity.

19. (Amended) A method as recited in Claim 1, further comprising the step of extracting similarity information from the similarity matrix to obtain new documents supported by the set of training documents for each category.

20. (Amended) A method as recited in Claim 19, wherein the similarity information is obtained by optimizing an objective function.

21. (Amended) A method as recited in Claim 19, wherein the similarity information is obtained by only approximately optimizing an objective function.

28. (Amended) A method as recited in Claim 1, further comprising the steps of creating and storing a second matrix using columns representing documents and rows representing user sessions, and wherein values of elements of the second matrix represent interest in a document shown by a particular user in a particular session.

32
29. (Amended) A method as recited in Claim 1, further comprising the steps of creating and storing a matrix using columns representing user sessions and rows representing documents, and wherein values of elements of the second matrix represent interest in a document shown by a particular user in a particular session.

34. (Amended) A computer-readable medium carrying one or more sequences of instructions, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:

establishing a plurality of training sets, wherein each training set is associated with a category and includes training documents that have been classified as belonging to said associated category;

determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories by determining similarity between said each document and the documents that belong to the training set of said category; and

wherein the step of determining similarity is performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity.

35. (Cancel) A method of categorizing a plurality of new electronic documents for use in a hypertext search system, the method comprising the steps of:
creating and storing a set of categories for the documents;
creating and storing a matrix, in which rows and columns identify documents, and in which each element of the matrix stores a value that represents a similarity among a pair of documents associated with a row and column that intersect at the element;
deriving each matrix value by combining two or more measures of similarity that are obtained by analysis of the documents.

36. (Cancel) A method as recited in Claim 35, further comprising the steps of:
for each measure of document similarity, creating and storing a graph of links;
creating and storing a combined graph that combines the graphs and that represents a generalized similarity of the documents;
computing a generalized similarity value for a pair of documents based on the combined graph.

37. (Cancel) A method as recited in Claim 36, further comprising the steps of
classifying unclassified documents into category nodes of a taxonomy structure associated with the hypertext search system based on the generalized similarity value in combination with a comparison of a set of pre-classified training set of documents with a set of unclassified documents, to carry out classification.